Package ‘bama’

May 3, 2020

**Title**  High Dimensional Bayesian Mediation Analysis

**Version**  1.0.1

**URL**  https://github.com/umich-cphds/bama

**BugReports**  https://github.com/umich-cphds/bama/issues

**Description**  Perform mediation analysis in the presence of high-dimensional mediators based on the potential outcome framework. Bayesian Mediation Analysis (BAMA), developed by Song et al (2019) <doi:10.1111/biom.13189>, relies on two Bayesian sparse linear mixed models to simultaneously analyze a relatively large number of mediators for a continuous exposure and outcome assuming a small number of mediators are truly active. This sparsity assumption also allows the extension of univariate mediator analysis by casting the identification of active mediators as a variable selection problem and applying Bayesian methods with continuous shrinkage priors on the effects.

**License**  GPL-3

**Encoding**  UTF-8

**LazyData**  true

**RoxygenNote**  7.1.0

**LinkingTo**  Rcpp, RcppArmadillo

**Imports**  Rcpp, parallel

**Depends**  R (>= 3.5)

**Suggests**  knitr, rmarkdown

**VignetteBuilder**  knitr

**NeedsCompilation**  yes

**Author**  Alexander Rix [aut, cre],
            Yanyi Song [aut]

**Maintainer**  Alexander Rix <alexrix@umich.edu>

**Repository**  CRAN

**Date/Publication**  2020-05-02 22:00:05 UTC
bama

Bayesian Mediation Analysis

Description

bama is a Bayesian inference method that uses continuous shrinkage priors for high-dimensional Bayesian mediation analysis, developed by Song et al (2019). bama provides estimates for the regression coefficients as well as the posterior inclusion probability for ranking mediators.

Usage

bama(
  Y,
  A,
  M,
  C1,
  C2,
  beta.m,
  alpha.a,
  burnin,
  ndraws,
  weights = NULL,
  k = 2,
  lm0 = 1e-04,
  lm1 = 1,
  l = 1
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Length n numeric outcome vector</td>
</tr>
<tr>
<td>A</td>
<td>Length n numeric exposure vector</td>
</tr>
<tr>
<td>M</td>
<td>n x p numeric matrix of mediators of Y and A</td>
</tr>
<tr>
<td>C1</td>
<td>n x nc1 numeric matrix of extra covariates to include in the outcome model</td>
</tr>
<tr>
<td>C2</td>
<td>n x nc2 numeric matrix of extra covariates to include in the mediator model</td>
</tr>
</tbody>
</table>
\textbf{bama} uses two regression models for the two conditional relationships, $Y|A, M, C$ and $M|A, C$. For the outcome model, \texttt{bama} uses

$$Y = M \beta_M + A \beta_A + C \beta_C + \epsilon_Y$$

For the mediator model, \texttt{bama} uses the model

$$M = A \alpha_A + C \alpha_C + \epsilon_M$$

For high dimensional tractability, \texttt{bama} employs continuous Bayesian shrinkage priors to select mediators and makes the two following assumptions: First, it assumes that all the potential mediators contribute small effects in mediating the exposure-outcome relationship. Second, it assumes that only a small proportion of mediators exhibit large effects ("active" mediators). \texttt{bama} uses a Metropolis-Hastings within Gibbs MCMC to generate posterior samples from the model.

\textbf{Value}

\texttt{bama} returns an object of type "bama" with 12 elements:

\begin{itemize}
\item \texttt{beta.m} \texttt{ndraws} x \texttt{p} matrix containing outcome model mediator coefficients.
\item \texttt{r1} \texttt{ndraws} x \texttt{p} matrix indicating whether or not each \texttt{beta.m} belongs to the larger normal component (1) or smaller normal component (0).
\item \texttt{alpha.a} \texttt{ndraws} x \texttt{p} matrix containing the mediator model exposure coefficients.
\item \texttt{r3} \texttt{ndraws} x \texttt{p} matrix indicating whether or not each \texttt{alpha.a} belongs to the larger normal component (1) or smaller normal component (0).
\item \texttt{beta.a} Vector of length \texttt{ndraws} containing the \texttt{beta.a} coefficient.
\item \texttt{pi.m} Vector of length \texttt{ndraws} containing the proportion of non zero \texttt{beta.m} coefficients.
\item \texttt{pi.a} Vector of length \texttt{ndraws} containing the proportion of non zero \texttt{alpha.a} coefficients.
\item \texttt{sigma.m0} Vector of length \texttt{ndraws} containing the standard deviation of the smaller normal component for mediator-outcome coefficients (\texttt{beta.m}).
\item \texttt{sigma.m1} Vector of length \texttt{ndraws} containing standard deviation of the larger normal component for mediator-outcome coefficients (\texttt{beta.m}).
\end{itemize}
**sigma.ma0** Vector of length `ndraws` containing standard deviation of the smaller normal component for exposure-mediator coefficients (alpha.a).

**sigma.ma1** Vector of length `ndraws` containing standard deviation of the larger normal component for exposure-mediator coefficients (alpha.a).

**call** The R call that generated the output.

**Author(s)**

Alexander Rix

**References**


**Examples**

```r
library(bama)

Y <- bama.data$y
A <- bama.data$a

# grab the mediators from the example data.frame
M <- as.matrix(bama.data[, paste0("m", 1:100)], nrow(bama.data))

# We just include the intercept term in this example as we have no covariates
C1 <- matrix(1, 1000, 1)
C2 <- matrix(1, 1000, 1)
beta.m <- rep(0, 100)
alpha.a <- rep(0, 100)

set.seed(12345)
out <- bama(Y, A, M, C1, C2, beta.m, alpha.a, burnin = 1000, ndraws = 100)

# The package includes a function to summarise output from 'bama'
summary <- summary(out)
head(summary)
```

---

**bama.data**

*Synthetic example data for bama*

**Description**

Synthetic example data for bama

**Usage**

bama.data
Format

A data.frame with 1000 observations on 102 variables:

- **y** Numeric response variable.
- **a** Numeric exposure variable.
- **m[1-100]** Numeric mediator variables

Description

*fdr.bama* uses the permutation test to estimate the null PIP distribution for each mediator and determines a threshold (based off of the *fdr* parameter) for significance.

Usage

```r
fdr.bama(
  Y,
  A,
  M,
  C1,
  C2,
  beta.m,
  alpha.a,
  burnin,
  ndraws,
  weights = NULL,
  npermutations = 200,
  fdr = 0.1,
  k = 2,
  lm0 = 1e-04,
  lm1 = 1,
  l = 1,
  mc.cores = 1,
  type = "PSOCK"
)
```

Arguments

- **Y** Length *n* numeric outcome vector
- **A** Length *n* numeric exposure vector
- **M** *n x p* numeric matrix of mediators of *Y* and *A*
- **C1** *n x nc1* numeric matrix of extra covariates to include in the outcome model
- **C2** *n x nc2* numeric matrix of extra covariates to include in the mediator model
betam Length p numeric vector of initial betam in the outcome model
alpha.a Length p numeric vector of initial alpha.a in the mediator model
burnin Number of iterations to run the MCMC before sampling
ndraws Number of draws to take from MCMC after the burnin period
weights Length n numeric vector of weights
npermutations The number of permutations to generate while estimating the null pip distribution. Default is 200
fdr False discovery rate. Default is 0.1
k Shape parameter prior for inverse gamma. Default is 2.0
lm0 Scale parameter prior for inverse gamma for the small normal components. Default is 1e-4
lm1 Scale parameter prior for inverse gamma for the large normal components. Default is 1.0
l Scale parameter prior for the other inverse gamma distributions. Default is 1.0
mc.cores The number of cores to use while running fdr.bama. fdr.bama uses the parallel package for parallelization, so see that for more information. Default is 1 core
type Type of cluster to make when mc.cores > 1. See makeCluster in the parallel package for more details. Default is "PSOCK"

Details
TODO

Value
fdr.bama returns a object of type "fdr.bama" with 5 elements:

bama.out Output from the bama run.
pip.null A p x npermutations matrices containing the estimated null PIP distribution for each mediator.
threshold The cutoff significance threshold for each PIP controlling for the false discovery rate.
fdr The false discovery rate used to calculate threshold.
call The R call that generated the output.

Author(s)
Alexander Rix

References
print.bama

Examples

library(bama)

Y <- bama.data$y
A <- bama.data$a

# grab the mediators from the example data.frame
M <- as.matrix(bama.data[, paste0("m", 1:100)], nrow(bama.data))

# We just include the intercept term in this example as we have no covariates
C1 <- matrix(1, 1000, 1)
C2 <- matrix(1, 1000, 1)
beta.m <- rep(0, 100)
alpha.a <- rep(0, 100)

set.seed(12345)

out <- fdr.bama(Y, A, M, C1, C2, beta.m, alpha.a, burnin = 1000,
                ndraws = 100, npermutations = 10)

# The package includes a function to summarise output from 'fdr.bama'
summary(out)

print.bama

Printing bama objects

Description

Print a bama object.

Usage

## S3 method for class 'bama'
print(x, ...)

Arguments

x An object of class 'bama'.

... Additional arguments to pass to print.data.frame or summary.bama
print.fdr.bama  

Printing bama objects

Description

Print a bama object.

Usage

```r
## S3 method for class 'fdr.bama'
print(x, ...)
```

Arguments

- `x`: An object of class 'bama'.
- `...`: Additional arguments to pass to print.data.frame or summary.bama

summary.bama

Summarize objects of type “bama”

Description

summary.bama summarizes the 'beta.m' estimates from bama and generates an overall estimate, credible interval, and posterior inclusion probability.

Usage

```r
## S3 method for class 'bama'
summary(object, rank = F, ci = c(0.025, 0.975), ...)
```

Arguments

- `object`: An object of class "bama".
- `rank`: Whether or not to rank the output by posterior inclusion probability. Default is TRUE.
- `ci`: The credible interval to calculate. `ci` should be a length 2 numeric vector specifying the upper and lower bounds of the CI. By default, `ci = c(0.025,.975)`. Additional optional arguments to summary

Value

A data.frame with 4 elements. The beta.m estimates, the estimates’ credible interval (which by default is 95\% inclusion probability (pip) of each ’beta.m’).
summary.fdr.bama

Summarize objects of type "fdr.bama"

Description

`summary.fdr.bama` summarizes the `beta.m` estimates from `fdr.bama` and for each mediator generates an overall estimate, credible interval, posterior inclusion probability (PIP), and PIP threshold for significance controlling for the specified false discovery rate (FDR).

Usage

```r
## S3 method for class 'fdr.bama'
summary(object, rank = F, ci = c(0.025, 0.975), fdr = object$fdr, filter = T, ...)
```

Arguments

- `object`: An object of class "bama".
- `rank`: Whether or not to rank the output by posterior inclusion probability. Default is TRUE.
- `ci`: The credible interval to calculate. `ci` should be a length 2 numeric vector specifying the upper and lower bounds of the CI. By default, `ci = c(0.025, 0.975)`.
- `fdr`: False discovery rate. By default, it is set to whatever the `fdr` of `object` is. However, it can be changed to recalculate the PIP cutoff threshold.
- `filter`: Whether or not to filter out mediators with PIP less than the PIP threshold.
- `...`: Additional optional arguments to `summary`

Value

A data.frame with 4 elements. The `beta.m` estimates, the estimates' credible interval (which by default is 95\% inclusion probability (pip) of each `beta.m`).
Index

*Topic datasets
  bama.data, 4

bama, 2
bama.data, 4

fdr.bama, 5

print.bama, 7
print.fdr.bama, 8

summary.bama, 8
summary.fdr.bama, 9